

Desarrollo y validación de un instrumento en Español para evaluar el desempeño de docentes clínicos a través de las percepciones de sus estudiantes

MARCELA BITRAN^{1a}, BELTRÁN MENA², ARNOLDO RIQUELME^{1,3}, OSLANDO PADILLA^{4b}, IGNACIO SÁNCHEZ⁵, RODRIGO MORENO¹

An instrument in Spanish to evaluate the performance of clinical teachers by students

Background: The modernization of clinical teaching has called for the creation of faculty development programs, and the design of suitable instruments to evaluate clinical teachers' performance. **Aim:** To report the development and validation of an instrument in Spanish designed to measure the students' perceptions of their clinical teachers' performance and to provide them with feedback to improve their teaching practices. **Material and Methods:** In a process that included the active participation of authorities, professors in charge of courses and internships, clinical teachers, students and medical education experts, we developed a 30-item questionnaire called MEDUC30 to evaluate the performance of clinical teachers by their students. The internal validity was assessed by factor analysis of 5,214 evaluations of 265 teachers, gathered from 2004 to 2007. The reliability was measured with the Cronbach's alpha coefficient and the generalizability coefficient (g). **Results:** MEDUC30 had good content and construct validity. Its internal structure was compatible with four factors: patient-centered teaching, teaching skills, assessment skills and learning climate, and it proved to be consistent with the structure anticipated by the theory. The scores were highly reliable (Cronbach's alpha: 0.97); five evaluations per teacher were sufficient to reach a reliability coefficient (g) of 0.8. **Conclusions:** MEDUC30 is a valid, reliable and useful instrument to evaluate the performance of clinical teachers. To our knowledge, this is the first instrument in Spanish for which solid validity and reliability evidences have been reported. We hope that MEDUC30 will be used to improve medical education in Spanish-speaking medical schools, providing teachers a specific feedback upon which to improve their pedagogical practice, and authorities with valuable information for the assessment of their faculty.

(Rev Med Chile 2010; 138: 685-693).

Key words: Education, medical, undergraduate; Educational measurement; Questionnaires.

¹Centro de Educación Médica.

²Oficina Editorial.

³Departamento de Gastroenterología.

⁴Departamento de Salud Pública.

⁵Departamento de Pediatría, Escuela de Medicina, Facultad de Medicina.

Pontificia Universidad Católica de Chile. Santiago de Chile.

^aBioquímica, PhD en Farmacología.

^bEstadístico.

Trabajo financiado por el proyecto MECESUP PUC 008-2000.

Recibido el 12 de agosto de 2009, aceptado el 26 de mayo de 2010.

Correspondencia a: Marcela Bitran PhD

Fono: 02-3543811

E-mail: mbitran@med.puc.cl

La reforma curricular de la carrera de medicina, iniciada en 1993, significó un cambio profundo en la forma de hacer docencia clínica en nuestra Escuela^{1,2}. De un modelo centrado en el profesor, en la enseñanza por disciplina y en

la docencia hospitalaria, pasamos a un sistema que incorpora la mayor parte de las nuevas estrategias educacionales tales como: enseñanza centrada en el estudiante, aprendizaje basado en casos clínicos y seminarios de grupos pequeños, creación de

cursos integrados, internados optativos y aumento de la docencia ambulatoria^{3,4}.

Aunque la enseñanza “al lado de la cama del enfermo” sigue siendo fundamental, ya no es la única instancia de aprendizaje clínico; hoy en día los alumnos entran en contacto con el paciente y/o su mundo más tempranamente en el currículo: aprenden en salas-espejo entrevistando “pacientes entrenados o actores”, en centros de atención ambulatoria, en los domicilios particulares de los enfermos y –también– en las aulas, discutiendo sobre casos y problemas de pacientes reales⁴.

Con el progresivo reemplazo de clases masivas por actividades en grupos pequeños, el docente clínico ha tenido que incorporar nuevas habilidades como “escucha activa”, control de sesión, *feedback* o retroalimentación estructurada, etc. Para enfrentar los desafíos asociados a la profesionalización de la docencia, la Escuela de Medicina creó en el año 2000 un Diplomado en Educación Médica (del que hoy en día se han graduado más de 120 académicos) y comenzó una modernización del sistema de evaluación de sus docentes.

En este contexto, desarrollamos un instrumento para conocer la percepción de los estudiantes sobre el desempeño de sus docentes clínicos. El objetivo que guió este proceso fue entregar al docente clínico un *feedback* específico que le motivara a perfeccionarse.

Optamos por crear un instrumento propio ya que –a la sazón– no había instrumentos en español publicados en revistas de corriente principal (según multibúsqueda realizada en MEDLINE, PubMed, Web of Science, EBM, CINAHL, Alerta al conocimiento, Proquest Medical Library, Proquest Nursing Journals and Allied Health y Proquest Science Journals). Se usó como base teórica el modelo de la Universidad de Stanford^{5,6} que plantea que una docencia clínica de excelencia involucra 7 dominios: clima propicio al aprendizaje, objetivos claros, promoción de la comprensión por sobre la memorización, promoción del autoaprendizaje, control de la sesión, *feedback* (retroalimentación) y evaluación acorde a los objetivos.

El desarrollo del instrumento implicó un proceso de búsqueda de consenso que duró 7 meses, en el que participaron el Director de Pregrado, los profesores jefes de todos los cursos en que se realiza docencia clínica, los profesores jefes de los Internados, el Director del Centro de Formación de Docentes Clínicos y el Director de la Oficina

de Educación Médica.

Para conocer la validez y confiabilidad de las evaluaciones, buscamos diversas fuentes de evidencia. La *American Psychological and Educational Research Associations* ha definido 5 fuentes de evidencia de validez: 1) contenido; 2) proceso de respuesta; 3) estructura interna; 4) relación con otras variables y 5) consecuencias⁷. La validez de contenido y de estructura interna han sido tradicionalmente las más usadas en la evaluación de la docencia clínica⁸, y son las que examinamos en este estudio. La primera se refiere al contenido y formato del instrumento y pretende determinar si éste es consistente con el concepto que se pretende medir^{9,10}. La validez de estructura interna indica el grado en que los *ítems* individuales del cuestionario representan al “constructo” que interesa medir, y se determina mediante análisis factorial^{7,11}.

Para evaluar la consistencia o reproducibilidad de los puntajes¹² estimamos el coeficiente alfa de Cronbach¹³ e hicimos un análisis de generalización (G-study). Éste es particularmente importante en las evaluaciones de desempeño, pues permite conocer la composición de la varianza total y determinar el número mínimo de encuestas necesario para obtener resultados fiables¹². Ambos indicadores de confiabilidad (Cronbach y coeficiente de generalización) fueron estimados en este estudio.

El propósito de este artículo es comunicar el desarrollo y validación de MEDUC30, un instrumento diseñado para evaluar el desempeño de los docentes clínicos y para proporcionarles un *feedback* específico y útil acerca de sus prácticas docentes.

Métodos

Desarrollo del instrumento

Para contar con el respaldo amplio de los distintos actores de la docencia clínica, usamos una técnica basada en la teoría fundada¹⁴ empleando el método Delfi modificado¹⁵. Como punto de partida, confeccionamos un borrador de 46 *ítems*, en base a un análisis crítico de la literatura relativa a las cualidades de un buen docente clínico¹⁵⁻¹⁷. Este cuestionario se estructuró en 8 dominios: los 7 dominios propuestos por el modelo educacional de Stanford⁶ y un dominio adicional destinado a evaluar si la docencia se centró en pacientes reales y en sus condiciones bio-psico-sociales. Este

8^{avo} dominio se denominó “docencia basada en pacientes”.

El cuestionario inicial se sometió al análisis de un grupo de 17 personas, integrado por el Director de Pregrado, los profesores jefes de todos los cursos en que se realiza docencia clínica (n = 7), los profesores jefes de los Internados (n = 7), los coordinadores docentes de las unidades académicas asociadas (del Hospital de Urgencia Asistencia Pública y del Hospital Dr. Sótero del Río), el Director del Centro de Formación de Docentes Clínicos y el Director de la Oficina de Educación Médica. Estos docentes conformaron un panel estable que se reunió 4 veces durante 7 meses para perfeccionar el cuestionario en un proceso de análisis iterativo. Su función fue analizar los *ítems* del cuestionario en cuanto a su pertinencia, claridad y suficiencia (respecto de la calidad de la docencia clínica) y sugerir cambios (incluir, modificar o eliminar) de los *ítems* o de la estructura del cuestionario. Para aceptar cada versión del cuestionario, se requirió del consenso de los miembros del panel, una vez que éste hubiera llegado a un punto de redundancia a partir del cual no se generaran nuevas sugerencias.

A partir del borrador original se originó un cuestionario de 40 *ítems*, que se aplicó a estudiantes de pregrado de 3° a 7° año, a quienes se solicitó que evaluaran (en forma anónima) el desempeño de sus docentes clínicos al final de cada rotación (4 semanas promedio; 5 a 6 estudiantes por docente), respondiendo a las afirmaciones en una escala tipo Likert de 6 puntos. Las opciones 1 a 4 de la escala (1: casi nunca, 2: a veces, 3: con frecuencia y 4: casi siempre) sirvieron para conocer la percepción de los estudiantes del desempeño de sus docentes, mientras que la 5 (no aplicable) y la 6 (no entiendo), para conocer su opinión sobre la pertinencia y claridad de los *ítems*.

Los resultados de esta aplicación (238 encuestas sobre 65 docentes clínicos) se analizaron en términos de confiabilidad, correlación inter-*ítem* y apreciaciones de los estudiantes acerca de su claridad y pertinencia. En base a esta información se eliminaron 10 preguntas, y se modificó la redacción de otras 9 para hacerlas más específicas y claras. La nueva versión del cuestionario estuvo compuesta de 30 *ítems* y usó una escala con 4 puntos (1: casi nunca, 2: a veces, 3: con frecuencia y 4: casi siempre). El último *ítem* corresponde a una evaluación global (EG) que permite conocer con-

sistencia entre la percepción general y el promedio de los 29 *ítems* previos. Este cuestionario se aplicó en el año 2003 a estudiantes de 3° a 7° año (793 evaluaciones sobre 91 docentes) y mostró valores de Crombach > 0,88; buenas correlaciones inter-*ítem* (> 0,75) y una estructura factorial compuesta de múltiples dominios. Esta versión se consideró satisfactoria y fue denominada “MEDUC30”.

En este artículo presentamos el análisis de las evaluaciones recogidas con el MEDUC30 entre 2004 y 2007 (5.214 encuestas sobre 265 docentes clínicos), realizado para confirmar su validez y confiabilidad.

Análisis de validez

Para los análisis de validez nos ceñimos a las definiciones de *American Psychological and Educational Research Associations*. Según ellas, la *validez de contenido* se refiere al contenido y formato del instrumento y pretende determinar si éste es consistente con el concepto que se pretende medir⁷⁻⁹. Esta definición fue concretada por Beckman et al, en 3 indicadores, a saber: 1) que el desarrollo del instrumento considere la opinión de profesores con amplia experiencia en docencia clínica; 2) que el instrumento contenga *ítems* previamente usados en contextos similares, y 3) que su desarrollo esté basado en teorías educacionales establecidas¹⁰.

Por otra parte, la *validez de estructura interna* se refiere al grado en que los *ítems* individuales del cuestionario representan al “constructo” que interesa medir, y se determina mediante análisis factorial⁷. Esta fuente de validez se evaluó mediante un análisis factorial exploratorio con rotación “Varimax”¹¹ y la extracción de factores se hizo de acuerdo al criterio de Kaiser Gutmann^{19,20} que considera únicamente a aquellos factores que tienen un valor propio (*eigenvalue*) > 1.

Análisis de confiabilidad

Medimos la confiabilidad de MEDUC30 a través de 2 métodos: el coeficiente alfa de Cronbach¹³, y un estudio de generalización (usando el programa Genova; Crick JE and Brennan RL). Este último permite conocer la proporción de la varianza total asociada a cada componente, y determinar el número mínimo de encuestas necesario para obtener resultados fiables, y sus resultados son expresados mediante el coeficiente *g* con rangos de 0 a 1⁹. Los datos se analizaron en conjunto, sin separar por sexo del docente,

año calendario ni curso del alumno, dado que el análisis inicial mostró que estas variables no contribuyen significativamente a la varianza (datos no mostrados). Los demás análisis estadísticos se hicieron con el programa SPSS (versión 15.0.1, SPSS Inc., USA).

Aspectos éticos

La evaluación fue anónima, de modo de proteger los intereses de los alumnos y asegurar que no hubiera repercusiones indeseadas sobre ellos (p. ej. cambios en la actitud de sus docentes en función de las evaluaciones hechas por los estudiantes). El estudio contó con la autorización y patrocinio de las autoridades de la Escuela.

Durante los 2 primeros años de aplicación del MEDUC30 (2004 y 2005), las autoridades de la escuela conocieron los resultados grupales pero no los individuales. Esta medida tuvo por objeto evitar juicios prematuros que pudieran derivarse del análisis de un número de encuestas menor que el necesario para asegurar la confiabilidad estadística de los resultados. Este número suele oscilar entre 5 y 10 encuestas para instrumentos similares⁹, de ahí la necesidad de dar un tiempo para que cada docente acumulara al menos 5 encuestas. A partir del 2006, los resultados individuales fueron conocidos por el Director de la Escuela, el Director de Pre-grado y el Jefe del Departamento respectivo, además de cada docente en particular.

Resultados

El cuestionario y su estadística descriptiva se muestran en la Tabla 1. Como se indica en ella, el porcentaje de respuesta de los ítems (i.e. cociente entre el número de encuestas en que hubo una respuesta y el número total de encuestas) varió entre 88 y 100% (valores < 70% indican que el ítem no es suficientemente claro o no es pertinente²⁰). La Tabla 1 reporta también los valores promedio para cada ítem; estos variaron entre 3,07 y 3,88 (en una escala 1 a 4 puntos).

Análisis de validez

Por originarse de un proceso en que participó un panel de expertos en docencia clínica y por estar basado en un modelo teórico validado, MEDUC30 cumple con los indicadores fundamentales que definen a los instrumentos con alta

validez de contenido, Esto implica que hay una correspondencia entre el formato y contenido de los ítems del instrumento, y lo que éste procura medir; en nuestro caso, las conductas asociadas a una docencia clínica de excelencia.

Respecto de la *validez de estructura interna*, el análisis factorial indicó que MEDUC30 tiene una estructura de 4 factores, determinados según el criterio *eigenvalue* de los valores propios > 1 (Figura 1); estos factores dieron cuenta de 62,2% de la varianza total. La Tabla 2 ilustra la distribución de los 29 ítems en los 4 factores extraídos. El primero, llamado *docencia centrada en pacientes* (DCP), contiene los ítems 1 al 5. Al segundo factor, compuesto por 13 ítems (6-8, 13-22), lo llamamos *habilidades docentes* (HD). El tercer factor aglutinó 5 ítems (9,11,12,23 y 25) y fue bautizado *habilidades de evaluación* (HE). Por último, con 6 ítems (10,24,26-29), identificamos el factor *clima de aprendizaje* (CA).

La Tabla 2 permita apreciar también la relación entre los factores extraídos y las dimensiones teóricas (dominios) en base a las cuales se estructuró el instrumento. La concordancia entre el factor docencia centrada en pacientes y la dimensión teórica del mismo nombre fue perfecta. Lo mismo ocurrió con Clima de Aprendizaje (Tabla 2); ambos factores mantuvieron todos los ítems de las dimensiones homónimas. En *habilidades docentes* (HD) coincidieron los ítems de las dimensiones objetivos (O), promoción del aprendizaje autodirigido (PA), promoción de la comprensión (PC) y control de la sesión (CS). Por otra parte, los ítems de evaluación (EV) y *feedback* (FB) confluyeron en el dominio HE, con la excepción del ítem 10 y el 24, que "cruzaron" a clima de aprendizaje (Tabla 2).

Los resultados obtenidos con el MEDUC30 presentaron alta confiabilidad global (α de Cronbach: 0,97 de un máximo de 1) y por factor (α de Cronbach: 0,93, 0,88, 0,89 y 0,86 para DCP, HD, HE y CA, respectivamente).

El estudio de generalización indicó que el mayor porcentaje de la varianza (25%) está asociado a la variable docentes (d), seguido de la interacción docente x estudiante x factor (d x e x f), y de estudiante (e) (Tabla 3). Diez encuestas por docente probaron ser suficientes para obtener un coeficiente de confiabilidad (coeficiente g) de 0,85, aunque ya con 5 encuestas la confiabilidad fue buena (coeficiente g = 0,8).

Tabla 1. Ítems y estadística descriptiva del MEDUC30

n°	Ítem	% Respuesta	Promedio	EE
1	La docencia que efectuó el tutor estuvo basada en los pacientes que vió contigo	98	3,74	,60
2	El tutor dió relevancia a los aspectos socio-culturales de pacientes específicos	99	3,43	,80
3	El tutor dió relevancia a los aspectos psicológicos de pacientes específicos	99	3,42	,81
4	El tutor dió relevancia a los problemas éticos de pacientes específicos	98	3,31	,86
5	Planteó los problemas relacionados con el costo/beneficio de las decisiones relativas a sus pacientes	98	3,61	,68
6	Explicó los objetivos que debías lograr en la rotación	100	3,41	,87
7	Organizó el proceso de aprendizaje en torno a los objetivos	100	3,47	,81
8	Explicó la importancia de los aprendizajes para tu futuro desempeño clínico	100	3,62	,68
9	Evaluó el nivel de partida de tus habilidades al comienzo de la rotación	98	3,07	,99
10	Fue justo al evaluar tu desempeño como alumno	88	3,63	,68
11	Evaluó personalmente tus habilidades (p.ej. cómo efectuabas las preguntas en una anamnesis, cómo palpabas el abdomen)	97	3,26	,96
12	Siguió personalmente tu progreso en el aprendizaje de las habilidades	97	3,29	,90
13	Demostó explícitamente el razonamiento clínico que utilizó en la toma de decisiones	97	3,72	,61
14	Enfatizó la comprensión de los conceptos, más que la memorización de la información	99	3,71	,62
15	Usó preguntas abiertas, que te estimularan a pensar (aclarar, justificar, correlacionar, criticar, evaluar, analizar, interpretar o predecir)	99	3,69	,63
16	Te ayudó a entender claramente conceptos clínicos que te parecían difíciles	99	3,63	,68
17	Te estimuló a aprender más sobre tus pacientes	98	3,58	,71
18	Te orientó en el empleo de métodos apropiados de búsqueda de información	98	3,16	,95
19	Estimuló a los estudiantes a participar en discusiones grupales	99	3,61	,72
20	Usó eficientemente el tiempo durante la actividad docente	100	3,51	,80
21	Se preocupó de minimizar las interrupciones externas	100	3,49	,82
22	Fue puntual	100	3,54	,89
23	Te comunicó durante el proceso de aprendizaje, la apreciación de tu desempeño en relación a los objetivos	98	3,23	,97
24	En las ocasiones en que te entregó <i>feedback</i> , lo hizo con tacto, sin herirte	94	3,67	,71
25	Te hizo sugerencias específicas para mejorar tu desempeño	98	3,33	,91
26	Fue respetuoso contigo	100	3,88	,43
27	Demostó entusiasmo por su labor docente	100	3,78	,58
28	Creó un ambiente no intimidante que favoreciera el aprendizaje	100	3,76	,61
29	Mostró disposición a escuchar a sus alumnos	100	3,81	,55
30	Se desempeñó como un muy buen tutor	99	3,72	,62

El porcentaje de respuesta corresponde al cociente entre el número de encuestas en que hubo una respuesta y el número total de encuestas, calculado para cada ítem. El promedio y error estándar fueron calculados en base a 5.214 evaluaciones sobre 265 docentes.

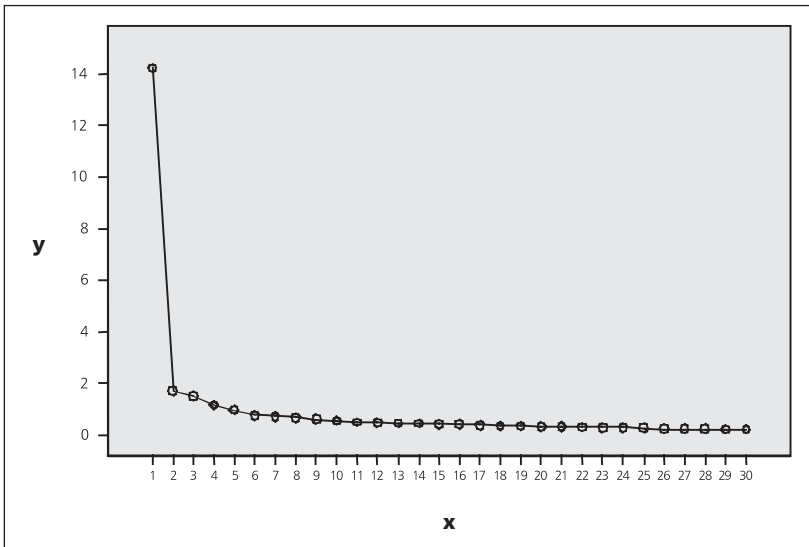


Figura 1. Gráfico de Sedimentación. El eje y consigna el valor propio (*eigenvalue*) en función del número de factores (eje x). El número de factores que compone un instrumento corresponde al valor de x cuando y es igual a 1; en este caso, el número de factores es un valor entre 4 y 5, y se aproxima a 4.

Tabla 2. Distribución de ítems del MEDUC30

Dimensión	Ítem	Factor			
		Docencia centrada en pacientes	Habilidades docentes	Habilidades de evaluación	Clima de aprendizaje
DCP	1	x			
	2	x			
	3	x			
	4	x			
	5	x			
O	6		x		
	7		x		
	8		x		
EV	9			x	
	10				x
	11			x	
	12			x	
PC	13		x		
	14		x		
	15		x		
	16		x		
PA	17		x		
	18		x		
	19		x		
CS	20		x		
	21		x		
	22		x		
FB	23			x	
	24				x
	25			x	
CA	26				x
	27				x
	28				x
	29				x

DCP: docencia centrada en pacientes, O: objetivos, EV: evaluación, PC: promoción de la comprensión, PA: promoción del autoaprendizaje, CS: control de la sesión, FB: *feedback*, CA: clima de aprendizaje.

Tabla 3. Estimación de los componentes de la varianza de las mediciones obtenidas con el MEDUC30 (a 5.214 evaluaciones sobre 265 docentes)

Fuente	componente de varianza	% de varianza
Docente (d)	0,10	25,1
Estudiante (e)	0,06	16,2
Factor (f)	0,04	10,7
Interacción docente-estudiante (d x e)	0,06	14,5
Interacción docente-factor (d x f)	0,03	8,8
Interacción estudiante-factor (e x f)	0,02	5,5
interacción docente-estudiante-factor (d x e x f)	0,08	19,1

Discusión

Comunicamos aquí el desarrollo y validación de MEDUC30, un instrumento para medir la percepción de los estudiantes sobre el desempeño de sus docentes clínicos.

Para juzgar la validez de las evaluaciones obtenidas con este instrumento, analizamos la *validez de contenido* y la *validez de estructura interna*, fuentes que han sido tradicionalmente las más usadas en la evaluación de la docencia clínica¹⁰. La primera se refiere al contenido y formato del instrumento y responde a la pregunta si el instrumento mide lo que pretende medir^{9,10}. La *validez de estructura interna*, por otra parte, indica el grado en que los ítems individuales del cuestionario representan al “constructo” que interesa medir. En otras palabras, si la “excelencia en docencia clínica” constara de múltiples dominios, la estructura de un instrumento diseñado para medirla debería componerse de múltiples factores, factibles de identificar mediante un análisis factorial^{7,11}.

Validez de contenido

MEDUC30 fue desarrollado para medir conductas propias de quien realiza una docencia clínica de excelencia. Podemos afirmar que MEDUC30 mide este concepto, es decir que tiene *validez de*

contenido, pues su desarrollo cumplió con los siguientes criterios¹⁰: 1) encuestar a profesores con experiencia respecto de la adecuación y representatividad de los ítems propuestos, 2) basarse en teorías educacionales establecidas y 3) contener ítems previamente usados en contextos similares¹⁰

Como se detalla en la sección Métodos, en la redacción y perfeccionamiento de los ítems del MEDUC30 participó un panel de expertos en docencia clínica (criterio 1), conformado por el Director de Pregrado, los profesores jefes de todos los cursos en que se realiza docencia clínica (n = 7), los profesores jefes de los Internados (n = 7), los coordinadores docentes de las unidades académicas asociadas (del Hospital de Urgencia Asistencia Pública y del Hospital Dr. Sótero del Río), el Director del Centro de Formación de Docentes Clínicos y el Director de la Oficina de Educación Médica. Por otra parte, el diseño del instrumento se basó en la teoría educacional de docencia clínica (criterio 2) desarrollada por la Universidad de Stanford y ampliamente aceptada^{5,6}. Por último, algunos de los ítems de MEDUC30 son similares a los del SFDP26, un instrumento de evaluación de docencia clínica altamente confiable y debidamente validado⁶ (criterio 3).

En consecuencia, por cumplir con todos estos criterios, podemos afirmar que MEDUC30 tiene *validez de concepto*. De hecho, se ubicaría en la categoría superior propuesta por Beckman et al, construida en base al análisis de 22 instrumentos conocidos de evaluación de la docencia clínica¹⁷.

Validez de estructura interna

La *validez de estructura interna* indica el grado en que los ítems individuales del cuestionario representan al “constructo” que interesa medir, y se determina mediante análisis factorial⁷. En el caso del MEDUC30, este análisis reveló la existencia de 4 factores básicos; dos de ellos conformados por los mismos ítems de las dimensiones teóricas homónimas (*docencia centrada en pacientes* [DCP] y *clima de aprendizaje* [CA]) y 2 factores compuestos (*habilidades docentes* (HD) y *habilidades de evaluación* (HE)), en los que confluyeron varias dimensiones. Este hallazgo confirma la naturaleza multifactorial de la “docencia clínica” y ofrece una versión simplificada del modelo teórico usado para construir el MEDUC30, que contempla 7 dimensiones^{5,6}. Apoya esta afirmación el hecho que, al “forzar” los datos experimentales a distribuirse en

8 grupos, los factores “compuestos” (HD y HE) se fragmentaron en las dimensiones teóricas aludidas (datos no mostrados). HD generó 4 sub-factores correspondientes a “objetivos”, “promoción de la comprensión”, “promoción del auto aprendizaje” y “control de la sesión”, y HE se dividió en 2: “evaluación” y “autoaprendizaje”.

En consecuencia, el análisis factorial aporta una evidencia sólida de la *validez de estructura interna* del MEDUC30. En el *ranking* propuesto por Beckman et al, nuestro instrumento se ubicaría en el límite superior que corresponde al criterio: “análisis consistente con la estructura anticipada”¹⁰.

Análisis de Confiabilidad

Las mediciones obtenidas con el MEDUC30 son altamente confiables; su coeficiente alfa 0,97 supera el criterio de exigencia para la toma de decisiones individuales y es notablemente mayor que 0,7, considerado aceptable para instrumentos nuevos²².

El estudio de generalización confirmó la confiabilidad de MEDUC30 indicando como principal fuente de varianza al docente. Este resultado es idéntico al reportado por Van der Hem-Stokroos²³ respecto del “*Clinical Teaching Effectiveness Instrument*” (CTEI). En ambos casos, la variabilidad asociada al docente es considerada una fuente deseable de varianza pues contribuye a discriminar entre individuos²².

Es sabido que la fiabilidad de los resultados varía con el número de cuestionarios²²; en el caso del MEDUC30, se necesitan al menos 5 encuestas por docente para asegurar una confiabilidad aceptable ($g = 0,80$), y 10 para una confiabilidad superior ($g = 0,85$ de un máximo de 1). Estos valores (5 a 10 encuestas) son fácilmente obtenibles dentro de un semestre académico, y son semejantes a los reportados para el *Clinical Teaching Effectiveness Instrument*²³.

Las fortalezas metodológicas del MEDUC30 residen en su validez, confiabilidad y en el tamaño de la base de datos en que se fundó el análisis. MEDUC30 puede compararse favorablemente con otros instrumentos utilizados en la evaluación de la docencia clínica en escuelas de medicina de Estados Unidos y Europa⁸. Además, tiene la enorme ventaja de estar en español. Este es un punto fundamental para las escuelas de medicina de la región, pues—en nuestro conocimiento—no existen instrumentos similares validados en español para

evaluar el desempeño del docente clínico.

Por haber sido desarrollado mediante un proceso participativo, MEDUC30 cuenta con la aceptación de la comunidad académica. Ésta es indispensable para promover la auto-superación y es considerada uno de los indicadores más importantes del potencial de impacto educativo de los instrumentos diseñados para evaluar competencia profesional²².

Como se indicó anteriormente, uno de las aspiraciones de este trabajo es que los resultados de la evaluación sirvan de *feedback* a los académicos para mejorar sus prácticas docentes. Esto está ocurriendo, a juzgar por los comentarios de docentes que afirman que las evaluaciones les han servido como guía para perfeccionar su desempeño docente, ya sea en forma autónoma, o tomando cursos del Diplomado de Educación Médica impartido por el Centro de Educación Médica de nuestra Escuela. Las evaluaciones de la docencia clínica tutorial no sólo son conocidas por los docentes sino también por los jefes de cursos, unidades docentes asociadas, jefes de departamento, y son en la actualidad un elemento más a considerar en el proceso de calificación y promoción académica.

Entre las limitaciones de este estudio debemos considerar que los mismos atributos que hacen del MEDUC30 un instrumento adecuado para la docencia clínica de nuestra Escuela, podrían comprometer su generalización a otros centros si sus ambientes educacionales son muy diferentes del nuestro. Sin embargo, nuestra opinión es que MEDUC30 puede ser de mayor utilidad para otras escuelas de medicina del país o de Latinoamérica que los instrumentos en inglés disponibles en la literatura. Además del aspecto idioma, comparadas con las escuelas de medicina de Estados Unidos o Europa, existen mayores semejanzas entre nuestras instituciones.

En todo caso, es indispensable validar este instrumento en la población específica en que se pretenda usar, pues, contrariamente a lo que se supone, la validez y confiabilidad son atributos de los puntajes y no de los instrumentos^{9,11}. De hecho, al interior de nuestra Escuela, hemos encontrado diferencias en la evaluación del desempeño de los docentes en los distintos campos clínicos (datos no mostrados).

Aunque las evaluaciones por estudiantes son un componente útil y valorado en la evaluación docente, no están exentas de sesgo²⁴. Por tanto,

la información proporcionada por el MEDUC30 debe ser complementada con otras fuentes, tales como autoevaluación y evaluación por pares y superiores jerárquicos. Esto es indispensable cuando se valora el desempeño global de un académico para la toma de decisiones importantes, como la promoción académica.

En conclusión, dada la rigurosa metodología utilizada en el desarrollo del MEDUC30 y sus sólidas propiedades psicométricas, consideramos que es un instrumento válido, confiable y útil para evaluar las percepciones de los alumnos respecto del desempeño de los docentes clínicos, particularmente para instituciones de habla hispana. MEDUC30 permite entregar a los docentes una retroalimentación específica de su desempeño y a las autoridades una información relevante para la valoración de sus académicos, ambas actividades indispensables para el perfeccionamiento de la educación médica.

Agradecimientos: Agradecemos a la Dra. Ximena Triviño y a la Ps. Denisse Zúñiga por la lectura crítica del manuscrito.

Referencias

1. Moreno R, Velasco N. [Curricular changes in the School of Medicine of the Pontificia Universidad Católica de Chile] *Rev Chil Cir* 1994; 46: 333-6.
2. Rosso P, Velasco N, Moreno R. [Undergraduate curriculum reform at the Pontifical Catholic University Medical School: aims, methodology and advance status]. *Rev Med Chile* 1997; 125: 796-807.
3. Harden RM, Sowden S, Dunn DR. Some educational strategies in curricular development: the SPICES model. ASME. Medical Education Research Booklet No. 18, 1984.
4. Sánchez I, Riquelme A, Moreno R, Mena B, Dagnino J, Grebe G. Revitalizing medical education: the school of medicine at the Pontificia Universidad Católica de Chile. *Clin Teach* 2008; 5: 57-61.
5. Skeff KM, Stratos GA, Berman MR. Improving clinical teaching: evaluation of a national dissemination program. *Arch Int Med* 1992; 152: 1156-61.
6. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998; 73: 688-95.
7. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington DC. American Educational Research Association, 1999.
8. Beckman TJ, Cook DA, Erwin PJ, Mandrekar, JN. How Reliable Are Assessments of Clinical Teaching? A review of published instruments. *J Int Med* 2004; 19: 971-7.
9. Downing S. Validity: on the meaningful interpretation of assessment data. *Med Educ* 2003; 37: 830-7.
10. Beckman T, Cook D, Mandrekar J. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005; 39: 904-10.
11. Kaiser H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958; 23: 187-200.
12. Downing S. Reliability: on the reproducibility of assessment data. *Med Educ* 2004; 38: 1006-12.
13. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
14. Glaser B, Strauss A. The discovery of grounded theory: strategies for qualitative research Chicago (IL): Aldine; 1967.
15. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nursing* 2000; 32: 1008-15.
16. Schmidt H, Moust J. What makes a tutor effective? A structural equations modeling approach to learning in problem-based curricula. *Acad Med* 1995; 70: 708-14
17. O' Malley P, Kroenke K, Ritter J, Dy N, Pangaro L. What learners and teachers value most in ambulatory educational encounters: a prospective, qualitative study. *Acad Med* 1999; 74: 186-90.
18. Irby D, Ramsey P, Gillmore G, Schaad D. Characteristics of effective clinical teachers of ambulatory care medicine. *Acad Med* 1991; 66: 54-5.
19. Field A. Discovering statistics using SPSS for Windows. Sage Publications Limited, London; 2000.
20. Field A. Discovering statistics using SPSS for Windows (Second Edition). Sage Publications Limited, London; 2005.
21. DeVellis RF. Scale development: Theory and applications. London, UK. Sage Publications, 1991.
22. van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Edu* 1996; 1: 41-67.
23. Van der Hem-Stokkoo H, Van der Vleuten C, Daelmans H, Haarman H, Scherpbier A. Reliability of the Clinical Teaching Effectiveness Instrument. *Med Educ* 2005; 30: 904-10.
24. Irby DM, Gillmore GM, Ramsey PG. Factors affecting ratings of clinical teachers by medical students and residents. *J Med Educ* 1987; 62: 1-7.